

DOCUMENT RESUME

ED 051 889

24

PS 004 833

AUTHOR Wright, Revilla; Levin, Barbara
 TITLE A Preschool Articulation and Language Screening for the Identification of Speech Disorders. Final Report.
 INSTITUTION Children's Hospital Medical Center of Northern California, Oakland.
 SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
 BUREAU NO BR-O-I-048
 PUB DATE Jun 71
 GRANT OEG-9-70-0016(057)
 NOTE 25p.
 EDRS PRICE EDRS Price MF-\$0.65 HC-\$3.29
 DESCRIPTORS *Articulation (Speech), Culture Free Tests, *Language Handicaps, Parents, *Preschool Children, School Personnel, *Screening Tests, Speech Handicaps, Standardized Tests, Tables (Data), Test Construction, *Test Reliability, Test Validity, Vocabulary

ABSTRACT

This report describes the validity and reliability studies done to standardize a preschool screening test developed for identification of articulation and language disorders. The test meets the need for a brief procedure which (1) identifies articulation and language disorders, (2) is not dialect bound and goes across cultural and regional differences in English, and (3) is standardized on a predominately black population. The test takes 5 to 8 minutes to administer and includes basic areas of vocabulary, articulation, and language. A failure in articulation and/or language indicates the need for a longer evaluation. The results of the short screening test are compared with the results of more extensive testing for 152 children enrolled in day care and Head Start type programs. Ages ranged from 2 1/2 to 5 1/2 years, with 25 children in each 6 month age group. Ninety-four percent of the time the screening either agreed with the evaluation (135 children) or gave a false positive (10 children), indicating that this test is an efficient and valid indicator of articulation and language disorders. Communication of results to school personnel and parents for effective follow up is discussed. Tables are included. A testing kit and manual are available. (Author/NH)

BR-0-I-048
PS

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
OFFICE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRO-
DUCED EXACTLY AS RECEIVED FROM
THE PERSON OR ORGANIZATION ORIG-
INATING IT. POINTS OF VIEW OR OPIN-
IONS STATED DO NOT NECESSARILY
REPRESENT OFFICIAL OFFICE OF EDU-
CATION POSITION OR POLICY.

ED051889

Final Report

Project No. O-I-048
Grant No. OEG-9-70-0016 (057)

A PRESCHOOL ARTICULATION AND LANGUAGE SCREENING FOR THE IDENTIFICATION OF SPEECH DISORDERS

Revilla Wright and Barbara Levin
Children's Hospital Medical Center of Northern California
51st and Grove Streets
Oakland, California 94609

June 1971

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF HEALTH, EDUCATION, AND WELFARE

Office of Education
(Educational Research)
Region IX

PS 004833

ACKNOWLEDGMENTS

One-hundred-fifty-two preschool children cooperated in this effort. Not only did they complete our screening, but responded to a long tedious pattern of tests. We remember their enthusiasm and willingness to help.

The administrators and personnel in the Preschool, Children's Centers, and Head Start programs of Oakland made the work possible. They participated actively and creatively with us throughout the study. They coped with the problem of fitting the testing into an already busy program. The children's cooperation with us is a direct result of their efforts. They prepared parents and children for our coming, and planned with us the follow-up program. They included: Mr. Sherman Skaggs, Director of Preschool and Children's Centers Programs of the Oakland Unified School District; Mrs. Beverly Ware, Director of Head Start, Oakland Economic Development Council; Mr. Carl Kennedy, Health Services Coordinator, Oakland Economic Development Council; Mrs. Fran Sherman, Director, Oakland Day Care Centers; Miss Lynn Pilchard and Mrs. Alfreda Wallace, social workers, Oakland Day Care Centers; Mrs. Carolyn Hunter, Preschool Program teacher on special assignment for curriculum; Mrs. Maxine Christopher, Consultant, Children's Centers; Mrs. Jerry Alcorn, teacher on special assignment for parent education; Mrs. Doris Bryan, Consultant for Health Services for the Oakland School District; Mrs. Margaret Outman, nurse for Preschool Program; Mrs. Carmen Anthony, teacher, North Oakland Head Start; Mrs. Ruth Lee, teacher, Golden Gate Preschool; Mrs. Valerie Ontiveras, teacher, Highland Preschool; Mrs. Shirley Jorgenson and Mrs. Arlene Davis, teacher, Lafayette Preschool; Mrs. Ella Averett, teacher, Washington Children's Center; Miss Anne Marie Morris, teacher, Golden Gate Children's Center; Mrs. Virginia Rose, teacher, Longfellow Children's Center; Mrs. Betty Foster, teacher, Clausen Children's Center; Mrs. Ora Taylor, teacher, Prescott Children's Center.

Children's Hospital Medical Center of Northern California provided the atmosphere in which such a project was possible. Mr. Harold Norman, Administrator, Dr. Edward Duffie, Jr., Medical Director, and Dr. Samuel Abraham, Director of Research, encouraged and supported the work. In addition, Dr. Duffie's expert counsel greatly helped us in the writing of our research proposal.

Dr. Louis Holm, Speech Pathologist, who volunteered his help in the early developmental stages and tested in the validation study, also was indispensable in the development of our proposal. Mrs. Gage Herman, Speech Pathologist, tested and evaluated spontaneous speech samples in the earlier stages of the screening development. Carol Beckland, Speech Pathologist, administered the screening; Mrs. Adeline McClatchie, Audiologist, and Mrs. Marie Ladd, Audiometrist, administered the hearing screenings. Nancy Redman, Speech Pathologist, played a major role in the validation study. She administered and rated tests, participated in follow-up conferences with teachers and aides, and was involved in planning and evaluation throughout the project.

Dr. Susan Ervin-Tripp, Ph. D., Associate Professor, University of California, Berkeley, applied her specialized knowledge and theory in psycholinguistics to the problem of evaluating language development in pre-schoolers. Dr. Maureen O'Sullivan, Ph. D., Psychologist and Statistician, Langley Porter Neuropsychiatric Institute and Assistant Clinical Professor of Medical Psychology, University of California Medical School, San Francisco, set up and interpreted our statistical procedures and also computerized our data. We also depended heavily upon her editorial abilities.

If oversights or errors were made they are the responsibility of the authors. If our test fulfills our hopes as an effective, useable screening device to identify speech problems in a preschool population, it is due to the exceptional talent of these people who have worked together in its development.

Revilla Wright
Barbara Levin

TABLE OF CONTENTS

I. Introduction.	1
II. Test Development and Pilot Studies	4
III. Validation Study.	7
IV. Implementation of the Screening Program	14
V. Conclusions	16
VI. References	17

LIST OF TABLES

1. Validation Study, 1970, Age Distribution by Sex, Racial and Ethnic Origin	18
2. Pass-Fail Results of Screening and Evaluation by Number of Children in Each Age Group	19
3. Correlations of Screening and Evaluation.	20
4. Internal Consistency Estimates from the Average Item-Test Correlation Corrected for Test Length.	21
5. Split Half Reliabilities Corrected for Test Length by the Spearman-Brown Extension Formula	21

INTRODUCTION

This report describes the validity and reliability studies done to standardize a preschool screening we had developed for identification of articulation and language disorders. It also includes descriptions of the test development, two pilot studies, and our implementation of the screening program. As background to this work, we here present a summary of the need for a preschool screening, the objectives of the test, some of the practical and theoretical considerations involved in its development, and the limitations of the screening test.

A speech screening for preschoolers is the first step of early intervention which could prevent long and costly remediation in later years and reduce secondary handicaps experienced by a child who has a serious communication problem. At the present time there are no standardized screening tests for the combined evaluation of articulation and language development which are applicable to the preschool child. The growth of preschool programs, such as Headstart and day care centers, has produced a great demand for articulation and language screening, and according to the Headstart manual, "Speech and hearing evaluations should be provided to all children, with referral for follow-up services where indicated".¹⁰ After an ERIC search through September, 1969, we found only one published study, by Monsee and Berman⁸ regarding speech screening at the preschool level for a Headstart population. In this report the researchers helpfully shared their experiences and pointed out the problems of attempting to score their test because items used in the test had no norms for the disadvantaged child.

Objectives: A standardized preschool screening test should meet the following conditions:

1. It should differentiate between children who have an articulation and/or language disorder, and those whose articulation and language development are within the normal range.
2. It should assess language, vocabulary, and articulation separately.
3. It should show the relationship among language, vocabulary, and articulation.
4. It should be short and interesting to the child. Disinterest and fatigue should be minimized.
5. It should be free of cultural bias. The test should detect speech problems per se, rather than cultural and regional differences in dialect and grammar.
6. Results should be communicated in such a way that effective and appropriate follow-up will be facilitated.

Practical or Theoretical Considerations Involved in Test Development:
Initial surprise at the absence of preschool screening procedures was

soon replaced by the realization of the difficulties involved. As we began work on establishing a screening it was necessary to deal with a number of inter-related problems: 1) Speech at the preschool age is going through a rapid period of growth and change, making it difficult to establish norms. Children tested at this age evidence a great deal of variation within the norm. 2) The test must be very short in order to be practical for large numbers of preschoolers. 3) Because a screening must be short, and because speech is changing rapidly at this age, there is the problem that once a test is reduced to screening length it will not distinguish between the child with a speech disorder and the child whose articulation and syntax are developing normally but at a somewhat slower-than-normal rate. 4) The preschool age child is often not yet using the adult articulation and language structure but a form that is appropriate for his age. Existing tests credit only the accepted terminal forms and give no credit for any others. Valuable information is lost in this approach as all the pre-adult forms used by a child are not examined or evaluated. We felt if we could distinguish between normal pre-adult forms and deviant or extremely delayed forms a screening could be short and also permit some differentiation between a child whose speech was developing at a somewhat slower-than-normal rate and the child whose speech was delayed or deviant. 5) To make this differentiation, it is necessary to know the developmental stages. The Poole Dissertation² investigated some early articulation substitutions and psycholinguists, such as our consultant, Dr. Susan Ervin-Tripp, have done extensive research on many aspects of language development. However, there are still no clear guidelines indicating the normal substitutions in vocabulary, articulation, and language before adult forms are reached. It was necessary to test a large number of items to find from our sample those items which most frequently indicated developmental stages. It was then possible for each item to have not just one score, but a range of scores. 6) Another concern was to construct a test that is not dialect bound, that would detect speech problems per se, not cultural or regional differences in articulation and grammar. The standardized tests in our field are based on Standard American English and a deviation is counted as an error. One consideration in selecting the articulation items was non-susceptibility to dialect. In the Sentence Imitation Section (language) responses in Standard and Nonstandard American English receive equal credit if they are of approximately the same complexity.

Limitations: This screening, to serve its function of testing large numbers of preschoolers, is extremely brief and for screening purposes only. Those using the test are cautioned that it is solely for the purpose of identifying children to be referred for an evaluation. The evaluation, and not the screening, determine if there indeed is a problem needing therapy.

Theoretically, the test should be as effective with the White middle class population as with the research population. Our work with private nursery schools and our cross-section clinic population indicate that this may be the case, but the screening has not been standardized on this group.

At present the screening is to be administered by speech pathologists. We have not yet explored the important possibility of the screening being administered by others, but we hope to do so.

SCREENING TEST DEVELOPMENT AND PILOT STUDIES

We began our screening test development in 1968 when we were asked to do a speech screening for five day care centers. From the longer procedures we devised the current screening procedures. In 1969 we did two pilot studies, one with children seen in our Speech and Hearing Center, and the other with children from five day care centers. The current validation study was done in 1970. The following describe each of these steps including description of the population, testing procedures, and testing results.

Test Development-1968: In 1968 we were asked to screen 82 children from five Oakland day care centers. These centers are federally funded by the Concentrated Employment Program and serve the children of parents who are in employment training programs or those of low income. Thirty-nine of the children were girls and 43 were boys. Seventy-four (90%) were Black, four (5%) were White, four (5%) were Mexican or other Latin Americans. Six children who spoke Spanish only were excluded from the sample. The children ranged in age from 2 years, 2 months, to 6 years 9 months. We have information on the family structure of 42 of the 82 children. Of these, 26% were living with their natural parents, 17% were living with one natural and one step-parent, and 57% were living with mother only.

Testing: Since no screening procedures were available, we used the following tests: 1) Children's Hospital Object Articulation Test, used to measure articulation; 2) Peabody Picture Vocabulary Test, Form B; 3) story telling from two series of pictures recorded and evaluated for language development and intelligibility; and 4) imitation of eight sentences of four to five words in length.

Results: Of the 82 children tested, 21 failed at least one section of the test. Thirteen children (16%) failed the articulation and/or language section of the testing. Seven children (8.5%) failed articulation, and eleven (13.4%) failed the language. Five children (6%) failed both articulation and language. Fourteen children (17%) failed the vocabulary section of the testing. Of those fourteen, six (7.3%) also failed articulation and/or language. Of the 21 children who failed at least one section of the testing, six were female and fifteen were male. Of the 13 children who failed articulation and/or language, eleven were male and only two were female. The six children who spoke only Spanish were excluded from the group because the test was not appropriate. The four bilingual Mexican-American children passed the test. Hearing screening was given at 20 dB (ISO) from 500 to 8000 Hz bilaterally. It was failed by two out of the 82, and one child would not participate in testing. Of these two, one passed the speech testing, and the other failed the naming section, but passed the articulation and language section.

PS 004833

The Screening Test: Our present screening test was based on analysis of responses obtained during the above testing. All of the items were devised by us and selected from a much larger number which we had tried initially. They are not part of an existing standardized procedure. The child is asked to name twelve objects, six for vocabulary, and six for articulation, and to repeat three sentences (later changed to five). Also included but not scored were check lists for peripheral speech mechanism and types of communicativeness and attention. The scoring is based on the norms of the day care center population tested earlier for the pilot study. Responses are scored and then scores are classified according to age groups as (1) pass, (2) borderline pass, and (3) fail. Only a fail in the articulation and/or sentence imitation (language) section indicates a need for a speech evaluation. Failing the vocabulary section does not indicate a speech disorder per se as it is more an indicator of general intelligence and cultural background, but it completes the pattern of the three basic language areas and gives the examiner a clearer picture of the nature of the problem. For example, a child with normal vocabulary who fails both articulation and language is likely to have a serious and often specific language problem. Special attention was given to item selection and scoring methods to make the test equally appropriate for children who speak either Standard or Nonstandard American English. The items chosen for the articulation section do not include the specific phonemes most likely to be affected by the dialects of the American Negroes. In scoring sentence imitation (language) responses in Standard and Nonstandard American English are given equal credit if they are of approximately the same complexity. The use of sentence imitation for the language section was the result of earlier work done by Nancy Wakefield, psycholinguist, and Revilla Wright.¹⁴ We had asked children in private nursery schools and in the Speech and Hearing Clinic to imitate sentences of increasing transformational difficulty, but of approximately the same length. We had used short sentences, and in general the children had no problem with them, making the study inconclusive. However, this result was in direct contrast to the speech clinic children with language problems who regularly had difficulty with the perception, processing, and speech production involved in imitation. Slobin or Welsh's¹¹ investigation of imitation as a device to learn about a child's language convinced them that sentence recognition and imitation are filtered through the individuals productive linguistic system. Our screening test does not include an assessment of receptive abilities. We are assuming a problem in reception or comprehension will be reflected in expression, and so for screening purposes testing expressive functioning will catch both. We are usually aware of the rare exceptions, such as a child who has suffered brain damage after having learned to talk, without a screening. In an evaluation, of course, separate assessment of comprehension and expression is essential.

Speech and Hearing Clinic Pilot Study-1969: Once the screening test was

devised, we evaluated it on 22 children ranging in age from 2 years, 6 months, to 5 years, 6 months who were referred to the Children's Hospital Speech and Hearing Center, Oakland, California, for articulation and/or language problems. Fifteen were boys, nine were girls, eleven were Black, and thirteen were White. The children represented a cross section of socio-economic classes in our area. Both the longer speech evaluation and the screening showed that 20 of the 22 children had a speech disorder. Two children passed the screening: one was a child who had been in therapy and was being terminated; the other was found to have a slight deviation in articulation (borderline pass) with normal vocabulary and language. This child had just been referred and the full evaluation showed identical results. In terms of potential validity, these results were encouraging.

Day Care Center Pilot Study-1969: A second pilot study of the screening procedure was done in the summer of 1969. Fifty-four children from the five Oakland day care centers who had not been tested previously were given the screening. All had normal hearing as determined by hearing screening at 20 dB (ISO) from 500 to 8000 Hz. Initially the speech screening was checked for inter-examiner reliability (percentage agreement across items=93.5%). The speech screening of 54 children was followed by more comprehensive testing of 23 of these children. The more comprehensive testing was done on: 1) children who failed any section of the screening, 2) children whose speech was questioned by the teacher, and 3) a random sample of those who passed the screening. Since these children came from the same day care centers seen in 1968, we assume that the racial or ethnic composition of the group was similar.

Testing: The results on the following tests were used as criteria to assess the validity of the short screening. For vocabulary we used the Peabody Picture Vocabulary Test; for articulation the Templin-Darley 50 Item Screening Test of Articulation plus 44 additional phonemes (elicited by naming objects); and for language five subtests of the Illinois Test of Psycholinguistic Abilities (visual reception, auditory association, auditory sequential memory, visual association, and grammatic closure), and the spontaneous speech section of the Houston Test for Language Development which judges communicative behavior, temporal content, syntactical complexity, and sentence length. An examination of the structure and function of the peripheral speech mechanism was also included. For those younger children who did not attend to the Illinois Test of Psycholinguistic Abilities and the Templin-Darley, we substituted the Houston and our object articulation test. To control for criterion contamination, those doing the full testing had no knowledge of the screening results. We compared the results of the longer testing on this small group of 23 with the screening results. The screening in each instance was as effective as the full testing in identifying children with articulation and language problems. It also identified children with minor speech deviations (i. e. pass but slightly below norm). By our method of selection some age groups were not included in the testing.

VALIDATION STUDY 1970

The validation study was done on 152 children from two-and-a-half to five-and-a-half years of age with 25 children in each six month age group. In each school where testing was done all children in the age range studied were seen until 25 had been tested for a six month age group, and from then on that age group was omitted. This continued until 25 children had been seen in each of the six age groups. Of the 82 girls and 70 boys, there were 142 Black, six White, three Mexican, and one Filipino. In Table I age distribution by both sex and racial and ethnic origin is reported. The children were enrolled in Head Start, the Oakland Public Schools' Children's Centers, or the Oakland Public Schools' Preschool Program for children whose families received low income or Welfare assistance.

To assess the effectiveness of the screening procedure in identifying children with speech disorders, each child was given our short screening by a speech pathologist and then given a longer speech evaluation by another speech pathologist. To avoid contamination there was no communication of results until all the scoring was completed and each child was rated: 1) Pass, 2) Borderline pass, or 3) Fail on the screening and on the evaluation. The results of the longer speech evaluation were used as the criteria for validating our short screening procedure.

Selection of Criteria: In selecting tests for the longer evaluation we were unable to find any which were not standardized on the White population. After investigating possibilities we made our selections on the basis of standardization and extensive use, and because we were testing preschoolers, on the basis of interest and length. In the language section we used only those tests that directly measured some aspect of language and omitted tests such as visual-association and perception. Although these selections were inadequate in some respects, we found similar or greater problems in the use of other measures. Since no one test adequately measures all three areas of vocabulary, articulation, and language, separate tests were used for each section. The Peabody Picture Vocabulary Test was standardized on the White population and one may question whether it adequately samples the vocabulary of the children we were testing. The Peabody also tests vocabulary comprehension and we were sampling expressive vocabulary. The Templin-Darley Test of Articulation¹³ was standardized on a White population and we did not know to what extent differences in dialect might affect the score. However, we assumed (correctly as it turned out) that the effect would be minimal on this age group. The language tests were the most obviously inadequate for our purposes. The Illinois Test of Psycholinguistic Abilities and the Houston Test for Language Development were standardized on a White population, the Northwestern Syntax Screening test on middle and upper-middle income communities and from homes where

Standard American dialect was spoken. In the sections involving grammar, credit was not given for mastery of the language system used in the environment unless it was Standard English. The Houston gives a more gross estimate in that scores are categorized by one year age intervals rather than the six month intervals that we use. By using clinical judgment as well as scores we work^e to minimize the effect of regional and cultural differences in language.

Procedure: The longer speech evaluation consisted mainly of the standardized tests just discussed and took about one and one-half hours to administer. It was given in two parts to the younger children.

Vocabulary: The Peabody Picture Vocabulary Test was used to assess vocabulary. We assigned a failure rating (3) to all vocabulary quotient scores of 65 and below, a borderline pass rating (2) to vocabulary quotients between 66 and 89, and a pass rating (1) to a vocabulary quotient of 90 or above. The child's rating on the Peabody was compared to his rating (1), (2), or (3) on the Naming section (vocabulary) of the screening.

Articulation: To test articulation we devised a 95 item test. In order to compare our screening results with a standardized procedure, we included the 50 items in the Templin-Darley Screening Test of Articulation. We felt the complete 128 picture items of the Templin-Darley were too long for children this age. However, because the screening form tested mainly later developing sounds, for our own information (and not scored for the Validation Study) we included items to test the earlier developing sounds in three positions (initial, medial, and final). To maintain interest, toy objects as well as pictures were used. We assigned a failure rating (3) to those children whose score fell on or below the Templin-Darley Screening Test cutoff score, which separates adequate from inadequate performance; a (2) rating to scores above the cutoff but below the norm; and a (1) rating to those at the norm or above. These ratings were compared with the child's rating on the articulation section of the screening. Since the Templin norms start at age three, we used a score of (1) as the cutoff for the children between two years, six months and two years, eight months and a score of (6) as the cutoff for the children between two years nine months and two years 11 months.

The longer evaluation for language was particularly difficult as we felt no one test was appropriate. We gave three subtests of the Illinois Test of Psycholinguistic Abilities: auditory sequencing, auditory association, and grammatic closure. We gave a tentative (3) rating if the child's standard score was 26 or below, a (2) rating if it was between 27 and 30, and a (1) rating for scores above 30. We used the spontaneous speech section of the Houston Test for Language Development which included

communicative behavior, temporal content, syntactical complexity, and sentence length. If a child's performance was more than two years below the norm in any of the areas, he was considered a tentative (3) rating. We initially tried the full Northwestern Syntax Screening with 20 pairs of Receptive items and 20 pairs of Expressive items. Because of problems of interest and inappropriateness for this population we soon reduced it to six pairs of Receptive sentences and nine pairs of Expressive sentences. There was no attempt to assign a rating. We also used some nonstandardized tasks. The children were asked to follow commands in two and three steps to assess their ability to follow a sequence of commands. Toys were used to examine comprehension of singular and plural nouns, both regular and irregular, and the possessive (example: Please give me the blocks, the children, the block, the child, etc., and, Please give me the mama's boy, the boy's daddy, etc.). Finally we devised a short illustrated story which the child was asked to re-tell after hearing it twice. This was recorded and later evaluated for intelligibility, syntactical complexity, and comprehension of the story. The speech pathologist evaluated the child's performance on the standardized and nonstandardized tasks and on the basis of clinical judgment made the final decision as to whether the child's language fell in the (1), (2), or (3) rating. That was essential as we found dialect often affected the language scores. The evaluation rating was compared with the rating on the Sentence Imitation section of the screening.

In addition to the three basic areas, the longer evaluation also included an evaluation of the structure and function of the peripheral speech mechanism, a Clinician's Impressions Check List to evaluate behavior during testing, and a hearing screening.*

*The hearing screening was given at 25 dB ISO (from 500 to 8000 Hz bilaterally) to 102 of the children. Due to illness, emergency surgery, etc. 50 children were not tested. Of the 102, seven did not participate in the testing, three of them two year olds. Of the remaining 95, 14 failed the hearing screening. One of these children failed articulation in both the speech evaluation and speech screening. The other 13 passed both speech evaluation and screening.

RESULTS OF VALIDATION STUDY

Validity: After the longer test results and the screening test results were scored and rated as (1) pass, (2) borderline pass, and (3) fail, the ratings on the longer test were compared with those on the screening. Ninety-four percent of the screening either agreed with the evaluation (135 children) or gave a false positive (10 children), i. e. screening failed but evaluation passed. The longer evaluations identified 32 children as having an articulation and/or language problem, i. e. a failure rating of (3). Of these, 25 were also given a (3) rating on the short screening. There were seven false negatives (4.6%), i. e. screening passed, but longer evaluation failed.

1970 VALIDATION STUDY N=152

Children failing screening and/or evaluation

Screening
failure



Evaluation
failure

110 or 72.4% children passed both screening and evaluation

The pass-fail results of screening and evaluation are given by number of children in each age group in Table II.

Discussion of False Positives and False Negatives: Seven children who passed the screening failed the longer testing (false negative). The longer testing indicates five had articulation problems and two had language problems. All five of the children who failed the articulation section of the longer procedure received a mental age of below three on the Peabody Picture Vocabulary Test (ranging from mental age two years, five months to two years, ten months) even though chronological age was somewhat higher. This is below the age range of the Templin-Darley which starts at three. With one exception these children correctly articulated all the early developing consonants which may indicate they were developing normally, but at a slower rate. Two children who passed the screening failed the language section of the longer testing. One was by default as the child remained silent for the activities involving conversational speech (Houston and story telling) giving the examiner no basis for assessing use of language. However, the child scored at the norm or above on the three Illinois Test of Psycholinguistic Abilities subtests used, so it may well be that the child had language at the norm or above and was not a failure. The other child failing the longer evaluation was in the fourth percentile on the Peabody Picture Vocabulary Test.

Although any screening needs to fail a higher percentage than a full evaluation, it should pass most of the children who have no problem. Of our group of 152, ten children who passed the longer evaluation failed the screening (false positives) which we felt was an acceptable error. Two of these failed articulation and eight failed the sentence imitation section. Both of these children were reported as having some articulation deviations in the longer evaluation, but did not fall in the fail category. Eight children, ranging in age from three years, three months to five years, two months, failed the sentence imitation section of the screening, but passed the language section of the longer evaluation. For all eight children, the evaluation examiner had checked that the child was "difficult to test" and that seven of the eight had "poor attention". The examiners also added the following descriptions: "negative", "couldn't get instructions", "some difficulty in understanding", "not interested-always wanted to leave", "very active - threw things", "difficult to follow instructions". One of these children refused to say the sentences and failed on that basis. Comparing this group of eight with the remaining overall group in this age range (3.0 years to 5.5 years) only 23% of the children were noted as being difficult to test as compared to 100% in the false positive group. The difficulty in testing observed by the examiner during the longer evaluation was not noted by the examiner of the short screening. However, it may be that without the screening examiner realizing it, problems with behavior

and attention were reflected in the child's performance on the sentence imitation. The manual now includes a caution with suggestions for reducing this effect.

Correlations: The scores of the articulation section of the screening have been correlated with the scores on the Templin-Darley 50 Item Screening and range from .608 (two-and-one-half year olds) to .869, with a mean of .765. The scores on the naming section of the screening were correlated with the scores on the Peabody Picture Vocabulary Test and these range from .356 to .595 with a mean of .463. These correlations are of moderate size and indicate that these measures are correlated although one is a measure of comprehension and the other of expression. The correlations for individual language tests with sentence imitation scores were variable, so we correlated the (1) pass, (2) borderline pass, and (3) fail from the examiner's overall language evaluation rating with the sentence imitation raw scores. These are negative correlations since we are correlating ratings where (1) is the highest and (3) the lowest with the screening scores where the higher the score the better the performance. These correlations range from -.355 to -.863. Table III shows the correlations for each of the areas: vocabulary, articulation, and language by age groups.

Reliability: Inter-examiner reliability was established first for the 1969 pilot study for the Day Care population. Two examiners scored separately screening of nine children. The percentage of agreement on the scoring for all items was 93.5%. The inter-examiner reliability was again assessed in 1970. This was done on a sample of 14 children with speech problems, not on the validation sample. One of the first examiners and one new examiner scored separately the screening of 14 children with speech problems between the ages of three years, four months and five years, five months. Out of the total number of 322 items there was scoring agreement for 298 items, an agreement of 92.5%.

For test re-test reliability twelve children between the ages of two years, six months and four years, three months were tested twice by the same examiner at an interval of 11 days. Out of a total number of 276 items there was scoring agreement for a total of 234 items or 85% agreement. The agreement by pass/fail scoring was 97%.

Internal Consistency: Two estimates of internal consistency were used. The first was estimated from the average item-test inter-correlations corrected for test length.* The internal consistency correlations for each section of the screening are given in Table IV. The internal

* Guilford, J. P., *Fundamental Statistics in Psychology and Education*, 4th Edition. McGraw-Hill, N. Y., 1965, p 463.

consistency correlations for naming (vocabulary) range from .551 to .789; for articulation the range was from .731 to .783; and for sentence imitation (language) the range was from .205 to .768. The reliability for these sentences is based on the last four of the five items. The second was estimated from the split half reliabilities corrected for test length by the Spearman-Brown formula. * These are given for each section of the test by age group in Table V. The correlations for naming (vocabulary) range from .299 to .720. For articulation the range was from .217 to .879. For sentence imitation (language) the range is from .601 to .801. The low correlations for the five year olds for naming and articulation are related to the restriction of range at that age. The brevity of the test mitigates against higher correlations.

The several discrepancies between the two internal consistency estimates may be due to the unreliability of reliability estimates based on small samples, or to sampling bias in the case of the split half estimates. The sentence imitation item-test reliability estimates may be lower because of the greater importance of attention in this test. Children's performances may be variable depending on the variability of their attention.

Changes Made in Screening Test After the Original Standardization: During the course of the validation study, scoring and response pattern difficulties suggested changes for the screening. These are as follows: Articulation-The articulation scoring originally gave only partial credit if a child used ~~tabah~~, tabo/table. However we found this response was common at all ages tested and that inter-examiner reliability was low in making the distinction between "tabah" and "table." The substitution of /ba/, /bo/, etc. for /bl/ is now given full credit and scoring cut-offs were adjusted to accommodate. This resulted in the following changes: two children now passed the articulation section of the screening; one who had failed the longer articulation test and one who had passed it. A third child who had previously passed now failed and this child had also failed the longer test. The effect was not reflected in the validation figures. Twenty-six children from age 3 1/2 and over were changed from borderline pass (2) to pass (1) category. Twenty-one of these had performed normally on the Templin. Sentences-Originally three sentences were used and we found scores clustering at the top for the older age groups. When doing the validation study, in addition to using the original three sentences, we had tried eight others. Two of these were selected to be included in the screening. These sentences were added to the original three and new scoring norms and cutoffs were established. Using the longer language evaluation as criterion, the additions to the sentence section resulted in four more false positives, and one additional true positive.

*Ibid. 1 458

IMPLEMENTATION OF THE SCREENING PROGRAM

The reason for a preschool speech screening program is to enable children with speech disorders to obtain help at an early age. As pointed out by the Bureau of Education for the Handicapped, "To take advantage of the period during which the most rapid growth and development takes place, work with handicapped children should start at the earliest feasible time. Positive intervention at this time will also diminish the negative effects of inappropriate learning and the accompanying frustrations experienced by young handicapped children and their parents".¹ An important prerequisite to early intervention is the involvement of teachers, health personnel, and parents in the screening program. Those doing screening programs are well aware that identification without follow-up is meaningless. However, this follow-up, far from being automatic, is a complex and often difficult process. It begins with initial contacts with school personnel and parents before rather than after the screening has been completed. Although the situations vary greatly from community to community (some have mobile units, others have public preschool services), we would like to share our experiences in the hope they may be useful to others doing screening programs in circumstances similar to ours.

As previously mentioned, the children we saw were mainly Black and from low income families. They were enrolled in programs which emphasize language development. For example, The language development activities for a typical day include speech production, language awareness, language patterns, vocabulary enrichment, auditory discrimination and cognitive development, as well as enrichment experiences and story groups.⁹ In our area the preschool programs do not provide speech therapy. Speech services are available through several clinics in the area where the children are seen on referral by a physician. These facilities have fees which are on a sliding scale or which may be covered by Medi-Cal, Children and Youth Project, or Crippled Childrens Services.

Our screening program included three main stages. Initial planning meetings with administrators and teachers and in some instances with parents, administration of the screening, and the follow-up. One purpose of the initial meetings with teachers and administrators was to plan jointly the implementation of the screening program. Each school selected its own way of informing parents - a meeting with us, a letter, a posted notice, or an explanation by the teacher. Preliminary plans were made for follow-up. The initial meetings also provided the opportunity to clarify our role as speech pathologists. We found in a previous study that the need for speech therapy and the need for language enrichment were easily confused. This occurred even though screening was always done at the request of the school. Because there is great emphasis on language in the preschool program, it was easy for those not familiar with speech

pathology to misunderstand the purpose of speech screening. It is important to clarify that the purpose is not to identify the need for enrichment or help with grammar, but rather to identify speech disorders. Some also expected that the test results would be the basis for advice for their language development activities for particular children. In our first screening program we responded by giving our ideas on language development and speech therapy-type activities, which further confused the areas of language enrichment and speech therapy. We found in this study we were able to be used more appropriately and effectively when, at the beginning, we thoroughly clarified our role. This aided the staff in becoming attuned to speech disorders and resulted in support from some who were often distrustful of testing programs.

After the screening and testing were completed, the results were discussed with the school staff. The teachers and aides brought up questions about specific children, discussed the effect of a speech problem on the children in the school situation, and shared knowledge that would affect referral recommendations. This mainly related to children with a multiplicity of problems where priorities were involved. The actual follow-up was done in various ways depending on the staff's wishes. The teacher first discussed the findings of screenings and longer testing with the parents as she was closest to them. Teachers reported it was not always easy to do this. Some had difficulty answering parents' questions about speech and some were concerned that the parents might interpret a recommendation for speech help as a criticism. It proved helpful in communicating results to arrange a joint conference with parents, teacher, and examiner, or specific telephone times when the examiner would be available for questions from parents or teachers. Referrals were made to the most convenient and practical service for the family rather than to one facility. Our own facility, for example would have involved a long commute for some of the families. If a child with a speech disorder was going to kindergarten the following semester, the school was notified of the results of the testing. When a public health nurse was involved she proved invaluable in cutting through red tape and in facilitating actual enrollment in an appropriate program.

CONCLUSIONS

This study validates a preschool screening test we developed for identification of articulation and language disorders. The results of the short screening test were compared with the results of more extensive testing for 152 children enrolled in child care centers and Head Start type programs. Ages range from two-and-a-half to five-and-a-half, with 25 children in each six month age group. Ninety-four percent of the time the screening either agreed with the evaluation (135 children) or gave a false positive (10 children). We feel our procedure meets the need for a preschool screening test which: 1) differentiates between children who have an articulation and/or language disorder, and those whose articulation and language development are within the normal range, 2) Screens vocabulary, articulation, and language separately, 3) Shows the relationship among vocabulary, articulation, and language, and 4) Detects speech problems per se rather than cultural and regional differences in dialect and grammar.

The implementation of a screening program includes defining methods for communicating results to teachers and health personnel. Clarification of the speech pathologists role, involvement of school personnel and parents from the initial planning stages, and a close working relationship with teachers and parents during follow-up all facilitated early intervention.

A testing kit and manual for administration and interpretation have been prepared.

REFERENCES

1. Bureau of Education for the Handicapped. Policies and Procedures. "Handicapped Children's Early Education Assistance Act". U. S. Office of Education, 1969.
2. Blanchard, I., The Genetic Development of the Articulation of Consonant Sounds: The Poole Dissertation, La Verne, California: Preston Printing, 1966 (revision).
3. Crabtree, M., The Houston Test for Language Development and Manual, Houston: 10133 Bassoon, 1958.
4. Dunn, L. M., Expanded Manual Peabody Picture Vocabulary Test, Minneapolis: American Guidance Service, 1965.
5. Guilford, J. P., Fundamental Statistics in Psychology and Education, 4th Ed. New York: McGraw Hill, 1965.
6. Lee, L., Northwestern Syntax Screening Test. Evanston: Northwestern University, 1969.
7. McCarthy, J. J., and Kirk, S. A., "The Construction, Standardization, and Statistical Characteristics of the Illinois Test of Psycholinguistic Abilities." Urbana: Institute for Research on Exceptional Children, University of Illinois, 1961.
8. Monsee, E. & Berman, C. "Speech & Language Screening in a Summer Headstart Program," Journal of Speech & Hearing Disorders, 33, 121-126, 1968.
9. Oakland Public Schools Research Dept. Evaluation Report, 1967-68 Unruh Bill Preschool Program, Research Report #3. Oakland: Oakland Public Schools, 1969.
10. Office of Economic Opportunity, Head Start; A Manual of Policies and Instructions, Washington, D. C.: Office of Economic Opportunity, 1967.
11. Slobin, Dan & Welsh, C., "Elicited Imitation as a Research Tool in Developmental Psycholinguistics, Working Paper #10. Berkeley: Language Behavior Research Laboratory, University of California, 1968.
12. Templin, M. C., Certain Language Skills in Children. Minneapolis: University of Minnesota Press, 1957.
13. Templin, M. C., and Darley, F. L., The Templin-Darley Tests of Articulation. Iowa City: Bureau of Educational Research and Service, Division of Educational Research and Service, Division of Extension and University Services, University of Iowa, 1960. (third printing 1965).
14. Van Riper, C., & Erickson, R., "A Predictive Screening Test of Articulation" Journal of Speech and Hearing Disorders, 34, 214-19 (1969).
15. Wakefield, Nancy, & Wright, Revilla, "Sentence Imitation Project" Unpublished manuscript. 1968.

Table 1

VALIDATION STUDY - 1970

Age Distribution by Sex, Racial, and Ethnic Origin

AGE GROUP	SEX		RACIAL/ETHNIC ORIGIN			TOTAL
	Boys	Girls	Black	Mex-Amer Latin-Amer.	White	
2.6-2.11	14	12	23		3	26
3.0-3.5	12	13	23	1	1	25
3.6-3.11	10	15	23		2	25
4.0-4.5	13	13	24	2*		26
4.6-4.11	10	15	24	1		25
5.0-5.5	14	11	24		1	25
Total	73	79	141	4	7	152
Percentage	48%	52%	93%	2 1/2	4 1/2	

*Filipino

Table 2

VALIDATION STUDY - 1970

Pass-Fail Results of Screening and Evaluation by Number of Children in Each Age Group

Overall Results

Age	Failed Scr. & Eval.	Failed Eval. Only	Failed Scr. Only	Passed Scr. & Eval.	Total
2.6-2.11	4	1	1	20	26
3.0-3.5	2	2	2	19	25
3.6-3.11	5	1	4	15	25
4.0-4.5	8	2	0	16	26
4.6-4.11	3	0	1	21	25
5.0-5.5	3	1	2	19	25
Total No.	25	7	10	110	152
Percentage	16.4%	4.6%	6.6%	72.4%	100%

Results by Sections

Articulation

Age	Failed Scr. & Eval.	Failed Eval. Only	Failed Scr. Only	Passed Scr. & Eval.	Total
2.6-2.11	2	1	1	22	26
3.0-3.5	1	3	0	21	25
3.6-3.11	4	1	0	20	25
4.0-4.5	6	2	0	18	26
4.6-4.11	3	0	1	21	25
5.0-5.5	2	1	1	21	25
Total No.	18	8	3	123	152
Percentage	12%	5%	2%	81%	100%

Language/Sentences

Age	Failed Scr. & Eval.	Failed Eval. Only	Failed Scr. Only	Passed Scr. & Eval.	Total
2.6-2.11	3	0	0	23	26
3.0-3.5	1	0	3	21	25
3.6-3.11	3	0	5	17	25
4.0-4.5	3	* 1	1	21	26
4.6-4.11	1	0	0	24	25
5.0-5.5	0	1	4	20	25
Total	11	2	13	126	152
Percentage	7.2%	1.3%	8.5%	83%	100%

Vocabulary/Naming

Results of this section not calculated in pass/fail results

Age	Failed Scr. & Eval.	Failed Eval. Only	Failed Scr. Only	Passed Scr. & Eval.	Total
2.6-2.11	0	0	1	25	26
3.0-3.5	0	0	1	24	25
3.6-3.11	1	2	2	20	25
4.0-4.5	3	1	2	20	26
4.6-4.11	0	2	4	19	25
5.0-5.5	0	0	2	23	25
Total	4	5	12	131	152
Percentage	2.6%	3.3%	7.9%	86.2%	100%

*No response

Table 3

CORRELATIONS OF SCREENING AND EVALUATION

Correlation of Peabody Score & Screening Naming (vocabulary) Score	Correlation of Templin Score & Screening Articulation Score	Correlation of Language Evaluation Rating & Screening Sentence Imitation (language) Score *
2.6-2.11 .477	.608	-.552
3.0-3.5 .595	.708	-.615
3.6-3.11 .256	.812	-.625
4.0-4.5 .463	.829	-.863
4.6-4.11 .346	.869	-.658
5.0-5.5 .302	.680	-.355
Total Group .414	.766	-.639
Ages 3-4 11 .424	.819	-.709

* These are negative correlations as in evaluation ratings (1) is the highest and (3) the lowest, these are correlated with screening scores.

Table 4

Internal Consistency Estimates From the Average Item-Test Correlation
Corrected for Test Length

AGE	Naming (vocabulary)	Articulation	Sentence Imitation (language)
2.6-2.11	.551	.749	.215
3.0-3.5	.756	.782	.482
3.6-3.11	.623	.832	.708
4.0-4.5	.745	.873	.205
4.6-4.11	.789	.852	.391
5.0-5.5	.678	.731	.768

Table V

Split Half Reliabilities Corrected for Test Length by the Spearman-Brown Extension Formula

Age	Naming (vocabulary)	Articulation	Sentence Imitation (language) *
2.6-2.11	.489	.835	.606
3.0-3.5	.680	.786	.775
3.6-3.11	.425	.825	.785
4.0-4.5	.469	.833	.800
4.6-4.11	.720	.879	.601
5.0-5.5	.299	.217	.696

*Reliability based on last four items of test.